DOCUMENT RESUME

ED 362 947 EA 025 281

AUTHOR Ligon, Glynn
TITLE Continuing Chapter 1's Leadership in Modeling Best Practices in Evaluation. A Symposium Presentation.
PUB DATE Apr 93
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
AVAILABLE FROM Evaluation Software Publishing, Inc., 3405 Glenview Ave., Austin, TX 78703.
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; *Educational Assessment; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; *Evaluation Problems; Measurement Techniques; Performance; Program Evaluation; Public Schools; School Based Management
IDENTIFIERS *Education Consolidation Improvement Act Chapter 1

ABSTRACT

This paper examines whether the Title I/Chapter 1 tradition of leading the way in educational evaluation will continue or whether Chapter 1 will change its role by delegating decision-making authority over evaluation methodology to state and local school systems. Whatever direction Chapter 1 takes, states, school systems, and schools must be held accountable for their activities. Conclusions are as follows: (1) Chapter 1 must define the purpose of its assessment; (2) Chapter 1 must select a methodology and instrumentation to answer that question; (3) more than one methodology and instrument may be needed to answer more than one question; (4) Chapter 1 must continue to mandate accountability and fund it; and (5) Chapter 1 must continue to develop and test evaluation methodology. Seven recommendations are offered regarding the design and implementation of accountability-focused evaluation. One figure is included. (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 362 947

Continuing Chapter 1's Leadership in Modeling Best Practices in Evaluation

A Symposium Presentation
by Glynn Ligon

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Ligon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented
at the 1993 Annual Meeting
of the American Educational Research Association
Atlanta, Georgia

EA 025281

Continuing Chapter 1's Leadership in Modeling Best Practices in Evaluation

Did Chapter 1 Lead?

Yes. As a participant/observer in the history of the development and maturation of program evaluation in public schools, I can appreciate the influence, even leadership, that was provided by Chapter 1, formerly Title I. A turning point was in the late sixties and early seventies when money, yes real money, became available to provide evaluation resources to meet accountability reporting requirements in large federal grants coming out of the original Title I of the Elementary and Secondary Education Act (ESEA), the Emergency School Aid Act (ESAA), and other substantial programs. During this period, public schools began to establish formal research and evaluation offices and the evaluation methodologies developed in universities were adopted by the school systems. This started the ongoing process of accommodating these evaluation models to the realities of public schools. During these times, it was not unusual for the research and evaluation office to be funded by Title I with other externally funded evaluations, such as for ESAA or Title VII Bilingual, along side. Title I set the pace because of the resources provided and the mandates imposed. The fact that Title I also published the TIERS evaluation models was important, but the critical contribution I see from Title I at that time was the issuing of a mandate for evaluation and the provision of resources to meet that mandate. Later came other requirements such as the use of NCEs for reporting of norm-referenced achievement test results for aggregation at the state and national levels. The provision of methodologies was helpful to move along the evolution of best practice as to how programs should be evaluated. Along the way, technical support centers were established to offer advice and assistance. All this created a most unusual environment where a mandate was imposed, methodology defined, funding provided, and a support system established.

Will Chapter 1 be able to answer the accountability questions Congress is asking?

The question I would raise in this discussion is whether the Title I/Chapter 1 tradition of leading the way will continue, or whether Chapter 1 will respond to recent trends and change its role by delegating to states and local school systems too much decision making authority over evaluation methodology. Too much delegation would be evidenced by there being so much variation across states and schools that interpreting the results at a national level would be difficult. Then the next question is: Will Chapter 1 be able to answer the accountability questions Congress is asking? Please, do not assume that Chapter 1 can answer these questions now. The limitations with NCEs and the quality of reporting from the states would need to be addressed even if no other changes were made to the regulations.

...control of the overall accountability measures must remain an organization-level responsibility.

Site-Based Management

If we look at the movement toward site-based management, we see that the idea is to move decision making and control over quality down to the levels where the people are who really know what works and where those people can make the changes necessary to improve quality. However, control of the overall accountability measures must remain an organization-level responsibility. Organizations must maintain a clear mission and central goals while decentralizing the authority to determine how to achieve those goals. This does not mean that schools and programs cannot develop or select their own measures of progress; not at all. What this means is that individual schools and programs must participate in the organization's accountability system as their way of demonstrating their contribution to the overall mission and goals of the organization. The message here is that whatever direction Chapter 1 takes,

The message here is that whatever direction Chapter 1 takes, it must not lose the ability to hold states, school systems, and schools accountable for their activities.

it must not lose the ability to hold states, school systems, and schools accountable for their activities.

...who determines the criteria for accountability for the millions of Chapter 1 dollars being spent.

Defining the Testing Debate

I would suggest that the debate with Chapter 1 and local schools regarding testing is not merely over the selection of a test, or an assessment system, but is a more fundamental debate of who determines the mission and goals of Chapter 1, and who determines the criteria for accountability for the millions of Chapter 1 dollars being spent. If Chapter 1 at the national level loses its ability to measure the effectiveness of the overall program by delegating control of the evaluation to states or school systems, then we will have not only distributed decision making, but have distributed accountability. Is that bad? Depending upon the questions we want to be able to answer about the effectiveness of our Chapter 1 dollars, that can been seen as bad or good. If we want to answer the question "Are individual states, school systems, and schools effective in meeting their objectives?" then this can be good. If we want to answer the question "Are individual states, school systems, and schools effective in contributing to the national goals and objectives as established by Chapter 1?" or if we want to answer the question "Which states, school systems, and schools are the most effective?," then this is not good.

A National Outcomes Evaluation

I would strongly urge that if accountability in Chapter 1 is to be decentralized, there be a national outcomes evaluation across the states that applies a standard measure of effectiveness on a sample basis. The sample could be large enough within each state to allow for a linking/equating of the state-adopted accountability measure with the national measure for comparisons within states. However, this equating would not be adequate to measure gains or to make gains comparisons using the individual state assessments.

Texas tried to squeeze added value out of its statewide criterion-referenced testing program, at that time called the Texas Educational Assessment of Minimum Skills (TEAMS) by equating scores with the Metropolitan Achievement Test (MAT6). Figure 1 illustrates what happened over time. Unfortunately, this was the same time period when Cannell was ridiculing states for all claiming to be above the national average, and Texas was inflating its statewide national percentile rankings through an artifact of equating a norm-referenced test and a criterion-referenced test. Simply, in 1985, TEAMS and MAT6 were equated; in 1988, that same equating table was used to place 1988 TEAMS scores on the MAT6 percentile scale. Because Texas school districts had concentrated their instructional efforts on the limited number of TEAMS skills, their TEAMS scores went up impressively, driving up the equated MAT6 percentiles. Realistically, the broader range of skills measured by MAT6 had not risen as much as the narrow range of basic skills measured by the TEAMS, so the equated MAT6 scores were artificially high. Texas abandoned this methodology when it was evident that predicting a score on a broader achievement test from a narrow criterion-referenced test is inappropriate over time.

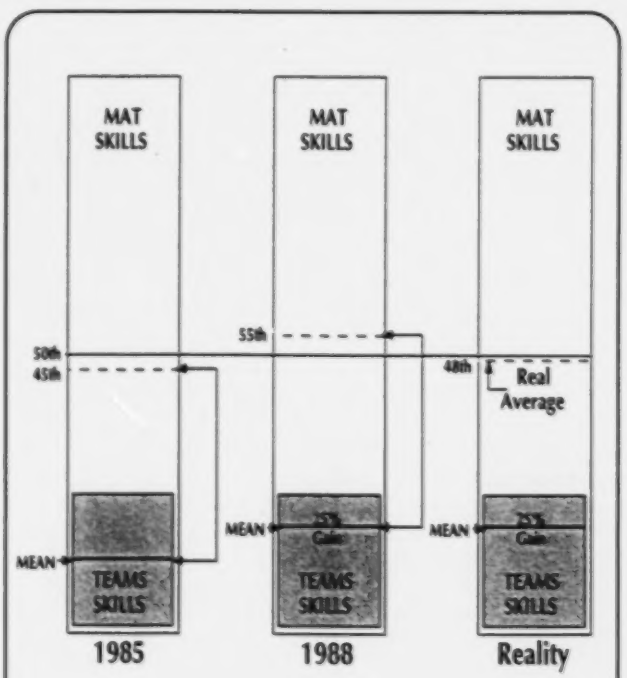

Figure 1

Continuing Chapter 1's Leadership in Modeling Best Practices in Evaluation

A National and International Perspective

You may insert here your own verbiage about how the United States is a mobile society that must recognize that individual schools are preparing students for national or even international competition, and that states must have comparative data beyond their own boarders to judge the effectiveness of their own programs. This issue was basic to the national education goals that emerged in 1989.

National Education Goals

The Education Summit of 1989 produced the National Education Goals Panel and six goals. A national Chapter 1 evaluation would be well served to address goals one through four, which are directly linked to elementary and secondary education. How each of these goals will be measured is still unsettled; however, the logic of linking Chapter 1 resources to these national goals is clear.

The National Education Goals

Goal 1: By the year 2000, all children in America will start school ready to learn.

Goal 2: By the year 2000, the high school graduation rate will increase to at least 90 percent.

Goal 3: By the year 2000, American students will leave grades 4, 8, and 12 having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so that they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.

Goal 4: By the year 2000, US students will be first in the world in science and mathematics achievement.

Goal 5: By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

Goal 6: By the year 2000, every school in America will be free of drugs and violence and will offer a disciplined environment conducive to learning.

Authentic Assessment Versus Traditional Tests

I would now like to fan the flames of an issue that many believe has been decided at the national level —the result of a battle conceded to the curriculum and instruction forces who consider norm-referenced, multiple-choice tests as the source of the high-stakes pressure to limit instruction to a narrow range of skills and content that can be measured by paper and pencil examinations. Meanwhile the advocates of accountability worry that performance assessments are so subjective that everyone will judge their programs successful while the nation's youth continue to achieve below international standards. I do not want to fight that battle here—merely to keep it alive. I do not want to concede that a final victory has been won by the authentic assessment advocates who are still working hard to produce an affordable, reliable, objective assessment system that is comparable across schools. Maybe we should stop here and examine that last phrase "comparable across schools." Many if not most authentic assessment advocates do not support comparisons across schools. That is a critical issue in the debate. Do we really need to compare the achievement gains of students in Austin, Texas, to the gains of students in New York to judge the effectiveness of Chapter 1 programs in each locality? Would we be satisfied if New York reported that based upon their standards as measured by their set of performance tasks, their Chapter 1 students were rated as more improved than comparable students who were not served? Would we be satisfied if Austin reported that its Chapter 1 students outperformed comparable low-achieving students on the Texas Assessment of Academic Skills (TAAS)? (By the way, you do not need to know what the TAAS is to answer this question, because the issue

...we should not rush away from the nationally normed, standardized, paper and pencil, achievement tests until we have available an alternative methodology that can answer Chapter 1's accountability questions.

is whether Texas should be able to use its own assessment measure as long as the state determines that it is appropriate.) Would we be satisfied knowing that Chapter 1 students in these localities performed better than their local peers, or would we want to be able to judge their performance based upon some external standard or norm? Is it satisfactory that Chapter 1 students within one state are outperforming other low-achieving students in that state without knowing whether those Chapter 1 students are falling behind or making up ground on students outside that state?

My point is that we should not rush away from the nationally normed, standardized, paper-and-pencil achievement tests until we have available an alternative methodology that can answer Chapter 1's accountability questions. If the authentic assessment advocates are still listening, then let me identify those accountability questions that are difficult to answer at this time outside the parameters of traditional assessment.

Accountability Questions

1. Does Chapter 1 funding result in greater success by disadvantaged students within the educational system than they would have achieved without those funds? (In simpler terms, is Chapter 1 service more effective than regular instruction?)
2. Is that higher success level adequate to remediate their educational disadvantage? (Is Chapter 1 making progress toward achieving its mission?)
3. Which individual Chapter 1 programs are successful and merit continuation and replication, and which programs are unsuccessful and require changes? (Where is the money being well spent and where is it not?)

Please note that none of these accountability questions addresses the issue of diagnosis and prescription of instructional activities. These are accountability questions, not the additional questions that a teacher, principal, program manager, instructional specialist, or program developer would need to ask. These are the questions asked by Congress, Chapter 1 staff in the Department of Education, school system administrators and trustees, taxpayers, and parents.

Strategic Planning for Assessment

Assessment and the resulting evaluations of programs are too important to be afterthoughts. Professional evaluators must influence their organizations to develop strategic plans for assessment. The goal of such a plan is to establish an information infrastructure that supports the data collection, analysis, and reporting systems required to provide information for decision making and management. A strategic plan for assessment would link a mission statement, goals, and objectives for an assessment program to the organization's mission, goals, and objectives. Then the organization's information audiences would be identified along with their information needs described in terms of the questions they need answered. From this, a plan would be developed to ensure that the information infrastructure and processes were in place. The benefit of such a strategic plan is that an organization is assured that answers to its critical questions are not dependent upon the changes that occur outside of their control. Changes such as those in state testing programs.

Chapter 1 on a national level is now struggling with the issues that would be addressed by a strategic plan for assessment. I trust that the final resolution of those issues will be based upon the questions that must be answered rather than upon other factors. Specifically, Chapter 1's plan for evaluation should be based upon the purpose for the assessments and the ability of those assessments to answer the three accountability questions identified earlier.

Let me quote from the testimony of Eleanore Chelimsky, Assistant Comptroller General, Program Evaluation and Methodology Division, U. S. Government Accounting Office, to the House of Representatives Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and Labor, February 18, 1993. Summarizing findings from ***Student Testing: Current Extent and Expen-***

> ***...it is not clear that one test can serve all three purposes...***
>
> ***Maximizing one purpose may degrade another...***

ditures, with Cost Estimates for a National Examination, she said:

"First, tension exists between our correspondents' preferences for two distinctly different emphases in testing: tests developed under local control and tests used principally for monitoring progress over time. Local control suggests a wide diversity of tests matched, in order to be most useful, to local variations in what is taught and learned; however, the goal of monitoring across classrooms, schools, districts, or states sets limits to the variation in tests that can be allowed without losing comparability. Second, tension exists between both local control and monitoring, on the one hand, and accountability, on the other. Although our respondents were not greatly concerned with accountability, others—chiefly outside the schools—have suggested that this purpose may be the most important; that is, using test results for high-stakes decision making about students, teachers or schools, and thereby emphasizing the importance of teaching and learning the material to be tested. Since it is not clear that one test can serve all three purposes, we conclude that decisions about test purposes are a high priority."

"...most issues of technical quality (for example, validity and reliability) and cost must be addressed in a specific context or purpose. Maximizing one purpose may degrade another: the research shows that the higher the stakes of a test, the more effort individuals will put into assuring high scores quite apart from genuine learning, which in turn makes the data less valid for monitoring. Our sense is that the debate over national tests has not yet distinguished clearly among the purposes to be served, nor has it drawn the appropriate conclusions concerning the technical difficulties involved in reconciling the conflicting requirements of a multipurpose test."

The importance of this testimony is that she understood from reviewing three reports on testing that the selection of a test/assessment must be driven, not by one's preferences for a certain type or one's bias as to what is authentic or standardized, but by one's purpose for testing. What a wonderfully simple notion—that we should select a test based upon the questions we are attempting to answer from the test results. In other words, if we begin with a strategic plan for assessment within which our questions are clearly defined, then we can select assessment instruments that match the purpose defined.

...we should select a test based upon the questions we are attempting to answer from the test results.

This points out the distinction that is missing from many debates on test types—different types of tests answer different types of questions; no single type of test answers all types of questions with the same degree of precision, reliability, and cost.

Practicality of Test Types for Different Purposes

The key issue may really be how practical a type of assessment is for the purpose intended. Both traditional and authentic types of measures might be developed, administered, and scored to be comparable across entities and reliable to an acceptable standard—the critical factor would then be the practicality of the methodology. Multiple-choice, paper and pencil tests have some distinct advantages: easy to administer, score, report; objective scoring; and with more creative item writing and scoring rubrics, ability to measure higher order thinking skills. They could be expanded to cover more areas and include more items in specific areas to address the curricular concerns that have been raised. Performance measures hold the advantage in being more valid in the sense that they are perceived as being closer to the behavior targeted for measurement. To achieve the same levels of reliability, performance measures have to include more tasks, involve more raters, and take more time than do traditional measures.

...if Chapter 1 is to conduct a national evaluation that is affordable and objective, then some form of traditional test makes sense at this point in time.

Practically speaking, if Chapter 1 is to conduct a national evaluation that is affordable and objective, then some form of traditional test makes sense at this point in time. If Chapter 1 develops its own test for this purpose, then the shortcomings of current off-the-shelf standardized tests can be addressed. However, the caution is that the measure needs to be an accountability tool, not a diagnostic tool. Chapter 1 should develop it to deliver on a single purpose—accountability.

The Education Economic Policy Center in Texas issued their recommendations for a statewide accountability system for Texas and surprised many by stating that the Norm-referenced Achievement Program for Texas (NAPT), the Iowa Tests of Basic Skills and the Tests of Achievement and Proficiency by their more recognized names, was the best available assessment instrument for the state. Their recommendation was based upon the charge that they had been given to design an accountability system. If they had been charged to design a diagnostic and prescription system, or even to ensure that their recommendation would produce diagnostic data useful for curricular analyses, they probably would have recommended the state's criterion referenced test. However, they were charged with recommending a measure that would serve the purpose of identifying the successful and unsuccessful schools—reliably, objectively, across time, and across grade levels.

Using the Same Test for Identification and Evaluation

Remember the debate about using the same measure for Chapter 1 identification and for a pretest? (The issue is whether selecting participants based upon a low pretest score provides an advantage in the pre-post gains from regression to the mean.) This is a similar debate—allowing the same measure that instructional staff want and need to diagnose and measure skill levels for placement to be used for accountability. Instructional placement does not require the objectivity that an accountability measure requires. Instructional placement can take advantage of a teacher's insight beyond the narrowly defined scoring criteria of either a performance measure or a paper and pencil test. Evaluation of Chapter 1 at the national level does not require the level of detail, the level of comprehensiveness, or the number of items that a diagnostic assessment does.

...performance assessment as defined in the authentic assessment movement will be the next great disappointment in public education.

I have said on more than one occasion that performance assessment as defined in the authentic assessment movement will be the next great disappointment in public education. I do believe that performance assessments will evolve into useful measurement tools that must become basic to good curriculum and instructional management and decision making. However, I do not believe that they will prove to be objective enough, free enough from bias, affordable on a large scale, or practical enough to double as our methodology for accountability.

So what does this all mean for Chapter 1? If Chapter 1 wants to maintain accountability to ensure that the money spent in their programs is having a positive impact on the learning of students to the extent that those students are making up the ground they are behind, then tests that answer the three accountability questions stated earlier must be mandated and administered. If Chapter 1 is satisfied with allowing local standards to prevail, with allowing subjective ratings to be compared across schools and states, then performance assessments are ready for endorsement and use for accountability.

Complaints about Testing and Tests

Listen carefully to many of the arguments against standardized tests and you hear some educators saying they do not want to be held accountable by those measures, but the alternative proposed is sometimes to establish a system where they rate their own students' performance and in a real sense judge themselves. I believe that as the authentic assessment movement matures, performance measures will lose their gloss as they are misused just as norm-referenced tests have been, and as performance measures become high stakes and are attacked for being reflective of only a part of the curriculum and being too unreliable to use as a basis for accountability.

Why do educators dislike NRTs so much? Because they, among other things:

- *receive too much attention,*
- *are used as the sole criterion or the most important criterion for decision making,*
- *are used for accountability, school/teacher/program evaluation,*
- *rank students and schools and lead to judgements about quality and ability and worth,*
- *do not cover everything taught,*
- *do not always reflect a student's true abilities/skills,*
- *take up instructional time,*
- *cost money, and*
- *focus instruction on a narrow range of content and skills.*

Eventually performance measures would inherit the same complaints. Add to this list the additional costs, instructional time, training, etc., and performance measures have a tough road ahead.

That is because many problems are not necessarily inherent in the tests, but in the use or misuse of them. However, please keep in mind that I believe and hope that authentic assessment is included in curriculum and instruction as best practice for teachers who are managing their instructional delivery based upon the learning of their students.

...problems are not necessarily inherent in the tests, but in the use or misuse of them.

Proposal

I would like to propose that Chapter 1 lead the way again by recognizing that assessment is a multitrack proposition. There are two main purposes for which assessments are needed in Chapter 1—diagnosis and accountability; different types of measurements are needed for each.

Diagnosis may require testing of every student; however, accountability does not have to require testing of all students. If the level of accountability chosen by Chapter 1 is the state, then statewide sampling would be sufficient to hold a state's Chapter 1 program accountable for improving learning.

There are two main purposes for which assessments are needed in Chapter 1—diagnosis and accountability; and different types of measurements are needed for each.

Standards for a Chapter 1 Evaluation

Chapter 1 must establish its standards for an acceptable evaluation. These should include:

1. Gains should be measured in order to document improvement beyond both past performance and the influence of socio-economic factors on past performance.

2. Gains should be described adequately in order to determine if program participants are making progress sufficient to close the gap between them and higher performing students. A 30th percentile student who makes a 2 NCE gain in grade 5 is probably farther behind grade level than before.

3. The accountability measure must be comparative, with the criterion for success being that Chapter 1 students gain more than they would have without the program.

4. The accountability measure must be broad in scope rather than focused only upon the content/skills being taught in the program. This is important because of the inherent "supplement versus supplant" issue. If a program achieves tremendous gains on a limited focus measure that is sensitive to the specific area taught in a Chapter 1 program, the gain achieved may be at the expense of skills and knowledge in other areas—therefore, was there really a gain for the student in the long run? This is related to the issue of basic skills measurement versus higher order skills measurement, from which

Chapter 1 chose to require that programs measure beyond the narrow range of basic skills.

5. The accountability measure must be reliable, objective, and otherwise psychometrically acceptable.

6. The outcomes documented must be able to be linked to a measure of implementation to ensure (a.) that the period measured by the gains measures a period of implementation, and (b.) that a program actually was provided. The sanctions considered for an ineffective program might be quite different if the ineffectiveness was a consequence of nonimplementation (possibly poor management) or a consequence of ineffective interventions (possibly a poor program design).

7. At the national level, it must be possible to aggregate gains across states.

Progress Toward Graduation

Now having said all this, I want to propose a performance measure to add to Chapter 1's requirements—"progress toward graduation." Progress toward graduation is defined here as how close a student is to being on pace, in terms of age and grade level or credits earned, with the normal pace for students moving through the educational system—the pace that normally describes a student who will graduate rather than drop out. Indeed this measure is filled with local mores and folkways, is permeated with subjective criteria for promotion and retention, is greatly influenced by local standards for earning course credits, and is highly dependent upon whether educators are socially promoting students. However, this measure is fundamental to public education, fundamental to the mission of schools: given all the local standards and requirements to which all students are held accountable, are Chapter 1 students progressing at a pace that predicts they will graduate rather than drop out?

This measure should not be reported just for the grades served by Chapter 1. Local schools select the grades in which Chapter 1 resources will be spent, but their overall objective is to get those students across the stage at graduation. Therefore, the index that should be reported is a percentage of students who are on pace for graduation across the entire school system. This index can be charted/tracked across years, can be adjusted for the entry age of students into a system or Chapter 1 program, can be compared to students not being served, and can be compared to national standards or levels. Best of all, it is truly authentic, because it measures the success a school system is having achieving its mission.

The Management Challenge

Chapter 1 evaluation must also lead the way by progressing to the point that it drives program management, which in turn ensures ongoing program improvement. One systemic problem in public education today is the reality that most best-practice planning and evaluation processes are performed as a matter of mandate, because they are required externally, rather than because they play a basic role in the management of the ongoing activities of the organization or program. Think about this for a minute. School/program/campus improvement plans: are they written and measured because that is how principals, leadership teams, or Chapter 1 program managers plan, organize, implement, and monitor their ongoing activities? No. Too many are developed and printed when required, set aside during implementation, then pulled out at the end of the year to look back and see what edits are needed to print the next year's plan and to perform whatever required measurement of objectives is necessary. There is much talk about high-stakes testing and the driving of instruction by what is measured on the tests; however, the reality appears to be more a phenomenon of worrying about the test scores in the days before the administration rather than creating a planning, implementation, evaluation, improvement cycle that is informed by the test results.

Continuing Chapter 1's Leadership in Modeling Best Practices in Evaluation

Conclusions

My conclusions are very simple:

1. Chapter 1 must define the purpose (the question to be answered) of its assessment.
2. Chapter 1 must select a methodology and instrumentation to answer that question.
3. There may be more than one question, thus more than one methodology and instrument will be needed.
4. Chapter 1 must continue to mandate accountability and fund it.
5. Chapter 1 must continue to develop and test evaluation methodology.

Recommendations

1. Chapter 1 should conduct a national accountability-focused evaluation to answer these three questions.
   - *Does Chapter 1 funding result in greater success by disadvantaged students within the educational system than they would have achieved without those funds?*
   - *Is that higher success level adequate to remediate their educational disadvantage?*
   - *Which individual Chapter 1 programs are successful and merit continuation and replication, and which are unsuccessful and require change?*
2. State and local Chapter 1 programs should design and implement accountability-focused evaluations to answer the same questions.
3. State and local programs should also design and implement curriculum-based diagnostic assessments.
4. The national education goals should be included in Chapter 1 evaluations.
5. Pace toward graduation, as an index associated with Goal 2, to increase the high school graduation rate, should be a long-term outcome measure in Chapter 1 evaluations.
6. Chapter 1 programs should develop and use program/campus improvement plans as real working management plans.
7. Chapter 1 programs should conduct strategic planning for assessment to ensure that all questions being asked by audiences can be answered.

Copies of this presentation may be ordered from:

Glynn Ligon
Evaluation Software Publishing, Inc.
3405 Glenview Avenue
Austin, Texas 78703
(512) 458-8364